

Biostatistics, Univ. of Makerere
Power Law Distributions in Bioinformatics: A Technical
Tutorial

Timo Koski KTH Department of Mathematics

1 Introduction

Power-law tails have been observed in the distributions of the sizes of incomes, cities, internet files, biological taxa, and, after the sequencing of genomes, in (size) distributions of molecular parts (like protein families, and folds, and their functions, occurrence of pseudogenes) in nature.

The purpose of this lecture is to present the results in [22, 23, 24], and in [12, 13] to the effect that the state probabilities of well known birth-and-death processes provides a natural source of power law distributions. This was first realized by G.U. Yule¹, who established a model (a pure birth process) to explain the observed size distribution of genera with respect to the number of species [31].

Mathematical models yielding power laws have also been derived by Hill, Rouault, and Simon, see [7, 8, 9, 25, 27]. The work of Simon, to the extent it coincides with Yule's, is presented in detail in the sequel. In appendix D there is an outline of Hill's theory on the emergence of power laws. Further fundamental mathematical work on power laws is found in [15], and [27]. H. Kesten [15] points out the biological interpretation of the multiplicative model (c.f., example 1.7 below) he is studying.

1.1 Definition and Examples of Power Laws

A discrete probability mass function p_k has a power-law tail or is a **Power law**, if it holds that

$$p_k := P(X = k) \sim k^{-\gamma}, \quad \text{as } k \rightarrow \infty. \quad (1.1)$$

The notation $f(x) \sim g(x)$ is explained in appendix A. A probability density function can also have a power-law tail defined in an analogous manner.

Let us take a look at a few enlightening and important examples.

¹George Udny Yule (1871–1951), educated in engineering and physics, later reader in statistics in Cambridge,

<http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Yule.html>

Example 1.1 (Zipf’s Law (rank-frequency form)) We count the frequencies of occurrences of some N events (e.g., English words in today’s issue of *New Vision*). Then we determine the rank k of each event by the frequency of occurrence (the most frequent is number one and so on). Then, if we consider p_k as the frequency of a word of rank k , this is very likely found to be

$$p_k = c \cdot k^{-\gamma}, \tag{1.2}$$

where γ is close to one, and where c is the normalizing constant

$$c = 1 / \sum_{k=1}^N k^{-\gamma}.$$

The probability mass function in (1.2) is known as Zipf’s law, and is an empirical or experimental assertion, which seems to arise in many situations, but is not based on any theoretical model². The case with $\gamma = 2$ is known as **(Zipf-)Lotka’s Law**³ [19], and was found as a bibliometric law on the number of authors making k contributions.

Many want to talk about Zipf’s laws strictly in the **rank-frequency specific form**, i.e., when referring to properties of ranked frequencies. However, in biology (or taxonomy) there is a legitimate **generic-specific form**⁴ of Zipf’s law. This deals with the numbers of species per genus, as in [31]. Hill [8, 9] provides connections between these two forms using Hill’s model.

Troll and beim Graben [30] show that Zipf’s law for word rank statistics is in a certain sense (see [30]) equivalent to a power law of word frequencies.

When plotted, for $\log p_k$ with Zipf’s law in (1.2) becomes a straight line as a function of $\log k$, with negative slope given by the exponent γ .

■

Example 1.2 (Discrete Pareto’s Law or Zipf-Mandelbrot’s law) Benoit Mandelbrot⁵ is said to have formulated the problem of finding the probability

²George Kingsley Zipf (1902-1950), linguist and philologist at Harvard University, studied Chinese languages and the statistical properties of language.

³Alfred James Lotka (1880-1949), a mathematician and biologist, more known for the predator-prey model (Lotka-Volterra) of population dynamics.

⁴The terminology for this distinction is due to [8].

⁵(1924–), mathematician known for fractal geometry
<http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Mandelbrot.html>

mass function p_k such that

$$\sum_{k=1}^R p_k \log_2 k$$

is minimized, under constraint that the entropy

$$H = - \sum_{k=1}^R p_k \log_2 p_k$$

has a value given in advance. The interpretation is minimization of the mean cost of words in a natural language, given average information H per word, and the potential number of words R . Mandelbrot found that the solution is

$$p_k = c \cdot (k + q)^{-\gamma}, \quad (1.3)$$

where c is the pertinent normalization constant. An analogy to the Mandelbrot scheme is suggested in [21] that replaces 'words' with proteins, and thinks in terms of energy minimization in protein synthesis. The parameter q represents then the bias of the organism towards positive regulation, i.e., repression of gene expression. These interpretations will probably also be met with scepticism.

This law had been earlier found by the economist V. Pareto⁶, as a frequency of wealth as a function of income category (above a certain bottom level). In plain words this means: most success seems to migrate to those people or companies who are already popular.

It is found in [17] that all observed gene-expression levels (e.g., yeast cells) appear to follow (1.3) with the range of k depending on the size of the gene-expression library.

■

⁶*Nordisk familjbok, Tjugoförsta bandet*, UGGLEUPPLAGAN, 1915: ” **Pareto** [-tå] Vilfredo, italiensk-schweizisk nationalekonom, född 1848 i Paris, (d. 1923, förf.anm.), utbildades till ingenjör, men öfvergick så småningom till nationalekonomien, . . . , P. har tilldragit sig mycken uppmärksamhet genom sin med matematiska formler demonstrerade och af rikhaltiga statistiska uppgifter belysta teori om inkomstfördelningen mellan de olika samhällsmedlemmarna i skilda länder, en fördelning som mindre motsvara en egentlig pyramid än en sådan med konkava sidor och konvex bas, en toppsnäcka enligt P:s egen beskrivning.”

Example 1.3 (Continuous Pareto Distribution) A probability distribution for a continuous random variable with the density

$$f(x) = \begin{cases} \frac{\alpha q^\alpha}{x^{\alpha+1}} & x > q, \\ 0 & x \leq q, \end{cases} \quad (1.4)$$

where $q > 0$, $\alpha > 0$, is called a **Pareto density** with parameters q and α . The distribution function is thus

$$F(x) = \int_{-\infty}^x f(u)du = \begin{cases} 1 - \frac{q^\alpha}{x^\alpha} & x \geq q, \\ 0 & x \leq q. \end{cases} \quad (1.5)$$

■

Example 1.4 (Yule-Simon's Law) G.U. Yule [31] explained theoretically in the 1920's the observation that the number of genera (p_k) having k species is distributed approximately as in Zipf's law (in the generic-specific form). We shall establish Yule's model (the pure birth process) and his result in example 5.2 and in section 6 below. Yule obtained a special case of the following probability mass function, which was later given a more general treatment in [27]

$$p_k = \delta B(\delta + 1, k), k = 1, 2, \dots, \quad (1.6)$$

Here $\delta > 0$ is real, $B(\delta + 1, k)$ is the *Beta function*, i.e.,

$$B(x, y) = \int_0^1 u^{x-1} \cdot (1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad (1.7)$$

where $\Gamma(\cdot)$ is, for z with positive real part, the Euler gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (1.8)$$

It is established in appendix B (see (B.5)) that

$$p_k \sim \delta \Gamma(\delta + 1) \cdot \frac{1}{k^\delta}, \quad \text{as } k \rightarrow \infty. \quad (1.9)$$

■

Example 1.5 (A Bayesian Zipf's Law) By assigning a prior distribution for numbers of species belonging to a genus Hill [7] derived for the expectation $E[p_k]$ of the number of genera having k species the expression

$$E[p_k] = \frac{1}{k(k-1)},$$

which is a weak version of Zipf's law. Hill obtained also the Yule-Simon law in this weak sense. This is discussed in appendix D. ■

Example 1.6 (Exponential Growth Observed at a Random Time)

Let us consider the deterministic process of exponential growth, or

$$X(t) = e^{\mu t}, \quad \mu > 0.$$

We observe, or kill, the process at an exponentially distributed time $T \in \text{Exp}(\nu)$. Now, we compute the distribution of the state of the process at random time of observing, or at random age, $X(T)$.

$$\begin{aligned} P(X(T) \leq x) &= P(e^{\mu T} \leq x) = P\left(T \leq \frac{\log x}{\mu}\right) = \\ &= 1 - e^{-\frac{\nu}{\mu} \log x} = 1 - \left(\frac{1}{x}\right)^{\frac{\nu}{\mu}}. \end{aligned}$$

Hence the probability density function of $X(T)$ is

$$f_{X(T)}(x) = \frac{d}{dx} P(X(T) \leq x) = \begin{cases} \frac{\nu}{\mu} \left(\frac{1}{x}\right)^{\frac{\nu}{\mu}-1} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

This is a power law, which holds for all x . This simple example is found in [23], which also gives additional, substantial, examples of stochastic processes killed at a random time, where the state of the killed (or observed) process has power-law tail. The idea of killing at an exponential time will be applied again in section 6.2. ■

Example 1.7 (A Multiplicative Process) Harry Kesten⁷ considered in [15] the (matrix version) of the following process.

$$S_{t+1} = Q_t + M_t S_t, \tag{1.10}$$

where M_t and Q_t are sequences positive independent random variables. He showed that the solution of this process

$$\begin{aligned} S_{t+1} &= Q_t + M_t(Q_{t-1} + M_{t-1}S_{t-1}) = \dots = \\ &= Q_t + M_t Q_{t-1} + M_t M_{t-1} Q_{t-2} + \dots + M_t M_{t-1} \dots M_2 Q_1 + M_t M_{t-1} \dots M_1 S_0 \end{aligned}$$

has a power-law tail. More on the physics of this kind of processes ('multiplicative processes') is found in [28].

■

1.2 Scale-Free Property

Let ξ be a random variable with the Pareto distribution in example 1.3. Then we have for $b > a > q$ using (1.5) that

$$P(\xi > b \mid \xi > a) = \frac{1 - F(b)}{1 - F(a)} = \left(\frac{a}{b}\right)^\alpha.$$

The term *scale-free* is used of any distribution (discrete or continuous or mixed) that looks essentially the same when looked at any scale, or such that

$$P(\xi > b \mid \xi > a)$$

depends only on the ratio a/b , and not on the individual scales a and b . Zipf's law is scale-free in this sense.

Recently the scale-free property has been observed for the degree distribution of many networks, where it is associated with the so-called small world phenomenon⁸. Examples are the World Wide Web, and human web of

⁷Professor emeritus of mathematics at Cornell University, amongst other things the holder of 'Tage Erlanders gästprofessur' (Vetenskapsrådet) during the year 2002, <http://www.vr.se/omvr/ff/sida.jsp?resourceId=728>

⁸A small world network is a graph in which the distribution of connectivity is not confined to any scale and where every node can be reached from each other by a small number of steps.

sexual contacts ([18]) and many networks of interaction in molecular biology. The mathematics (or physics) of the scale-free property and power laws in networks is reviewed in [2, 3].

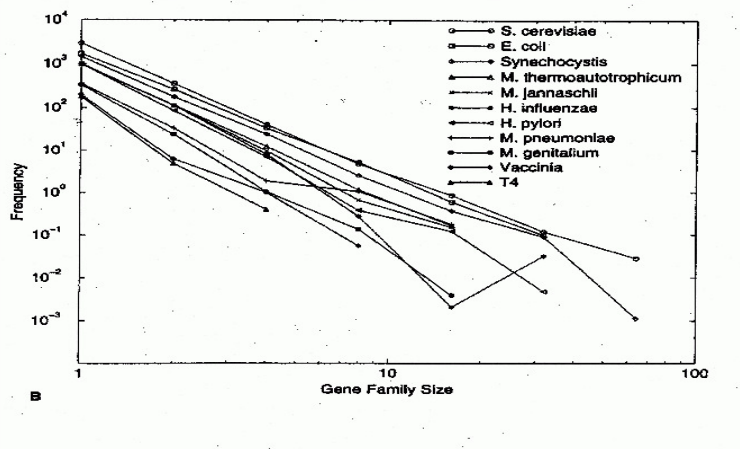
1.3 The Presence of Power Laws in Genomic Data

It has been found empirically (see, e.g., [11] and [20] and their references) that the power law behaviour applies to the distribution of a wide range of genome-associated quantities. These include the frequency distribution of gene family sizes, the number of transcripts per gene, the number of interactions per protein, the number of genes or pseudogenes in paralogous families, the occurrence of DNA words (short base sequences), as well as distributions of the connections of enzymes and metabolites in metabolic networks, the frequencies of distinct DNA and protein domains. Power laws have also emerged as distributions of lengths of biased random walks in non-coding DNA [1].

A figure digitally scanned from [11] shows the frequencies of gene family sizes. The family size is exponentially binned⁹. Linear regression shows that the slopes of the curves are

$$-2.18, 2.84, -3.17, -3.17, -3.27, -3.62, -3.45, -2.69, -4.02, -3.8,$$

and these slopes differ significantly from zero.



⁹Family size 1 falls in class 1, family sizes 2 and 3 in class 2, family sizes 4, 5, 6, and 7 fall in class 3, e.t.c., in logarithmic scale.

1.4 The Meaning of Power Laws in Genomic Data

The preceding list of observed cases of power laws could be dismissed as a mere list of biological trivia. However, as pointed out in [20], the power law behaviour is a mathematical-statistical expression of an important biological feature: *the dominance of a few members over the overall population*.

To quote one example of many in [20], out of the 247 distinct protein folds currently assigned in the worm genome, just 10 account for the over half of the 7805 assigned domains.

2 Birth and Death Processes: a General Case

One possible explanation of how the power laws arise, e.g., at the protein level, could be an underlying process of gene duplication. Should we treat gene duplication as a stochastic process, the chance of a given gene being duplicated is proportional to its occurrence in the genome. With each duplication some genes will increase their presence in the population, enhancing their chance of further duplication. Combined with selective pressure accounting for the functional significance of different protein products, such a process gives prominence to some gene types, or families, over others.

Next we describe a general class of stochastic processes with relevant properties, a general model for the branchings of a species into two species or extinction of a species. This is known as a birth-and-death process, which turns out to be linked to power laws in various ways.

2.1 Birth and Death Process: The Generator

A birth and death process is a continuous time Markov chain for the branchings of a species into two species or extinction of a species in which the rates of birth and death are constant through time. For the facts on continuous time Markov chains required here we refer to, e.g., [6, 16].

Lineages give rise to new lineages at rates λ_i , where i is the number of current lineages. At each split the parent can be thought of as either being replaced by its two daughters, or as adding a new member to the population and remaining a member itself, and the process moves from i to $i + 1$. The parameters λ_i represent the reproductive power of the population, effects of sex and age are ignored. Similarly, lineages go extinct at rates μ_i . A lineage

does not lose its power to reproduce before it dies. The reproductive and mortality effects are acting simultaneously and independently of each other.

We have the generator

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\mu_3 + \lambda_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \quad (2.11)$$

where

$$\mu_0 = 0, \lambda_0 \geq 0, \mu_i \geq 0, \lambda_i \geq 0, \quad i = 1, 2, \dots,$$

This generates a well-defined continuous time Markov chain $X = \{X(t) | t \geq 0\}$, which has the non-negative integers as state space.

Example 2.1 (Linear Birth-and Death Process) A *linear birth-and-death* process is obtained by

$$\mu_i = \mu i, \lambda_i = \lambda i, i \geq 1, \lambda_0 > 0. \quad (2.12)$$

■

Example 2.2 (Poisson Process) A *Poisson process* is obtained by

$$\mu_i = 0, \lambda_i = \lambda, i \geq 1, \lambda_0 = \lambda. \quad (2.13)$$

■

We know that, if $\mathbf{p}(t)$ is the row vector with components,

$$p_i(t) = P(X(t) = i),$$

then we have

$$\mathbf{p}'(t) = \mathbf{p}(t)\mathbf{Q}. \quad (2.14)$$

Due to (2.11) this can be written componentwise as (a system of differential-difference equations)

$$p_0'(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t) \quad (2.15)$$

$$p_j'(t) = \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t), \quad j \geq 1,$$

with some initial condition $\mathbf{p}(0)$.

2.2 Equilibrium Distributions for Birth and Death Processes

The equilibrium distribution, if it exists, satisfies the system of equations

$$\pi \mathbf{Q} = \mathbf{0},$$

where $\mathbf{0}$ is a vector of zeros. Hence we obtain from (2.15) the equations

$$0 = -\lambda_0 \pi_0 + \mu_1 \pi_1 \tag{2.16}$$

$$0 = \lambda_{j-1} \pi_{j-1} - (\lambda_j + \mu_j) \pi_j + \mu_{j+1} \pi_{j+1}, \quad j \geq 1.$$

We solve these by induction. We set

$$w_0 = 1, w_j = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}, \quad j \geq 1. \tag{2.17}$$

Then we get from (2.16) that

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0 = w_1 \pi_0.$$

If we assume that $\pi_k = w_k \pi_0$ for $k = 1, \dots, j$, we have

$$\begin{aligned} \mu_{j+1} \pi_{j+1} &= -\lambda_{j-1} \pi_{j-1} + (\lambda_j + \mu_j) \pi_j \\ &= -\lambda_{j-1} w_{j-1} \pi_0 + (\lambda_j + \mu_j) w_j \pi_0 \\ &= \lambda_j w_j \pi_0 + (\mu_j w_j - \lambda_{j-1} w_{j-1}) \pi_0. \end{aligned}$$

From the definition of w_j in (2.17) we get $\mu_j w_j - \lambda_{j-1} w_{j-1} = 0$ and thus we obtain

$$\mu_{j+1} \pi_{j+1} = \lambda_j w_j \pi_0 \Leftrightarrow \pi_{j+1} = \frac{\lambda_j}{\mu_{j+1}} w_j \pi_0 = w_{j+1} \pi_0.$$

This completes the induction. We see that

$$\sum_{j=0}^{\infty} \pi_j = 1 \Leftrightarrow 1 + \sum_{j=1}^{\infty} w_j < \infty. \tag{2.18}$$

is required for the existence of an equilibrium distribution.

Example 2.3 [*Equilibrium Distribution for a Linear Birth and Death Process*] For the linear birth and death process, where with regard to (2.18) we find

$$\sum_{j=1}^{\infty} w_j = \frac{\lambda_0}{\mu} \sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{\lambda}{\mu}\right)^{j-1} = -\frac{\lambda_0}{\lambda} \log\left(1 - \frac{\lambda}{\mu}\right)$$

under the assumption that

$$\frac{\lambda}{\mu} < 1.$$

This means clearly that the process has an equilibrium distribution, as soon as the death rate is larger than the birth rate for all $i \geq 1$. We set

$$C = \frac{1}{1 - \frac{\lambda_0}{\lambda} \log\left(1 - \frac{\lambda}{\mu}\right)}.$$

Then the equilibrium distribution for the linear birth and death process is

$$\pi_j = \begin{cases} C & j = 0, \\ C \frac{\lambda_0}{\mu} \frac{1}{j} \left(\frac{\lambda}{\mu}\right)^{j-1} & j \geq 1. \end{cases} \quad (2.19)$$

■

3 Birth and Death Processes for Frequencies of Domain Families

3.1 Another interpretation of birth-and-death

In [12] the following picture is painted. A genome is treated as a bag of genes coding for protein domains, to be called domains. Domains are treated as independently evolving units. Each domain is considered as a member of a family, which may have one or more members. Following events are considered:

- domain birth, which generates a new member in the same family as a result of gene duplication.
- domain death

- innovation, which generates a new family with one member. Innovation may occur via domain evolution from a non-coding sequence, or a sequence of non-globular protein, via horizontal gene transfer from another species or in other ways.

The rates of the elementary events are considered to be independent of time, and of structure, biological function, and other features of families. We assume that N is the maximum possible number of domain family members.

This is a situation similar to that Yule [31] considered in the 1920's. We study first a birth-and death process with a finite number of states and reflection at boundaries.

3.2 The Generator

We consider a birth-and-death process with a finite number of states and reflecting boundaries. Here

$$p_i(t) = \text{the relative frequency of a domain family of size } i, i = 0, \dots, N$$

$$= P(X(t) = i).$$

The forward equations for these probabilities are

$$\begin{aligned} p_0'(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ p_j'(t) &= \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t), \quad 1 < j < N, \\ p_N'(t) &= \lambda_{N-1} p_{N-1}(t) - \mu_N p_N(t), \end{aligned} \tag{3.20}$$

with some initial condition $\mathbf{p}(0)$. The generator is

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots & 0 \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & \dots & 0 \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \lambda_{N-2} & 0 \\ \vdots & \vdots & \vdots & \ddots & -(\mu_{N-1} + \lambda_{N-1}) & \lambda_{N-1} \\ 0 & 0 & 0 & \dots & \mu_N & -\mu_N \end{pmatrix}, \tag{3.21}$$

We set

$$C = \frac{1}{1 + \sum_{l=1}^N \prod_{i=1}^l \frac{\lambda_{i-1}}{\mu_i}}$$

This model has a unique equilibrium distribution given by

$$\pi_j = \begin{cases} C & j = 0 \\ C \prod_{i=1}^j \frac{\lambda_{i-1}}{\mu_i} & 1 \leq j \leq N. \end{cases} \quad (3.22)$$

A variant of the above describes, e.g., the evolution of the size of a domain family that includes an essential (indispensable) gene that is not allowed to go extinct, and is

$$p'_j(t) = \lambda_{j-1}p_{j-1}(t) - (\lambda_j + \mu_j)p_j(t) + \mu_{j+1}p_{j+1}(t), \quad 1 < j < N, \quad (3.23)$$

$$p'_N(t) = \lambda_{N-1}p_{N-1}(t) - \mu_N p_N(t), \quad j = N,$$

with some initial condition $\mathbf{p}(0)$. We set

$$C = \frac{1}{1 + \sum_{l=2}^N \prod_{i=2}^l \frac{\lambda_{i-1}}{\mu_i}}.$$

For this model the unique equilibrium distribution is given by

$$\pi_j = \begin{cases} C & j = 1 \\ C \prod_{i=2}^j \frac{\lambda_{i-1}}{\mu_i} & 2 \leq j \leq N. \end{cases} \quad (3.24)$$

4 Power Laws from Equilibrium Distributions of Birth and Death Processes

Let us assume that the birth-and-death process is linear, i.e.

$$\mu_i = \mu i, \lambda_i = \lambda i, i \geq 1, \lambda_0 > 0. \quad (4.25)$$

Then we get in (3.22)

$$\pi_j = C \prod_{i=1}^j \frac{\lambda_{i-1}}{\mu_i} \sim C \left(\frac{\lambda}{\mu} \right)^j \cdot \frac{1}{j}. \quad (4.26)$$

Hence, if $\lambda = \mu$, we have a power law

$$\pi_j \sim \frac{1}{j}, \quad (4.27)$$

i.e., the exponent γ in (1.1) is equal to one.

Let us consider a slightly more complicated model, where

$$\mu_i = \mu(i + a), \lambda_i = \lambda(i + a), i \geq 1, \lambda_0 > 0. \quad (4.28)$$

Then we get in (3.22)

$$\pi_j = C \prod_{i=1}^j \frac{\lambda_{i-1}}{\mu_i} = C \frac{\Gamma(1+b)}{\Gamma(1+a)} \frac{\lambda_0}{\lambda} \left(\frac{\lambda}{\mu}\right)^j \frac{\Gamma(j+a)}{\Gamma(j+1+b)}. \quad (4.29)$$

In view of (B.5) we get with $x = i + a$ and $y = 1 + b - a$

$$\frac{\Gamma(j+a)}{\Gamma(j+1+b)} \sim \frac{1}{j^{1+b-a}}$$

and thus

$$\pi_j \sim \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j^{1+b-a}}.$$

This probability goes, however, to zero faster than a power law due to the factor $\left(\frac{\lambda}{\mu}\right)^j$. This is a power law if $\lambda = \mu$. This analysis has been extended to more complicated forms of birth and death rates as functions of i in [13, 14]. Next we show another way power laws arise from birth-and-death processes.

5 Palm's Formulae for a Linear Birth and Death Process with Absorption at Zero

Let $\lambda_0 = 0$ so that zero is an absorbing state. The equations (2.15) reduce to

$$p_0'(t) = \mu p_1(t) \quad (5.30)$$

$$p_j'(t) = (j-1)\lambda p_{j-1}(t) - j(\lambda + \mu)p_j(t) + (j+1)\mu p_{j+1}(t), \quad j \geq 1.$$

The initial conditions are

$$p_1(0) = 1, \quad p_j(0) = 0, \quad \text{for } j \neq 1. \quad (5.31)$$

These equations describe the flow of probability in a Markov branching process, which starts at state 1, i.e., with one ancestor. We assume that

$$\lambda \neq \mu. \quad (5.32)$$

Let us introduce the auxiliary function

$$\beta(t) := \frac{1 - e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}} \quad (5.33)$$

which can be rewritten as

$$\beta(t) = \frac{e^{-(\lambda-\mu)t} - 1}{\mu e^{-(\lambda-\mu)t} - \lambda}, \quad (5.34)$$

which is well defined since $\mu \neq \lambda$ holds. Conny Palm¹⁰ solved the differential-difference equations (5.30) by the use of probability generating functions outlined in Appendix C. It holds that

$$p_0(t) = \mu\beta(t) = \frac{\mu e^{-(\lambda-\mu)t} - \mu}{\mu e^{-(\lambda-\mu)t} - \lambda}, \quad t > 0, \quad (5.35)$$

and

$$p_1(t) = (1 - p_0(t)) \cdot (1 - \lambda\beta(t)), \quad t > 0, \quad (5.36)$$

which yields

$$p_1(t) = \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^2}, \quad t > 0, \quad (5.37)$$

and

$$p_n(t) = p_1(t) \cdot (\lambda\beta(t))^{n-1}, \quad t > 0, \quad n \geq 1. \quad (5.38)$$

This is a *geometric distribution* with a modification for $n = 0$. We expand the expression by some simple algebra, and get

$$p_n(t) = \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^2} \left[\frac{\lambda e^{-(\lambda-\mu)t} - \lambda}{\mu e^{-(\lambda-\mu)t} - \lambda} \right]^{n-1}, \quad t > 0, \quad n \geq 1. \quad (5.39)$$

Since zero is absorbing, $p_0(t)$ is the *probability of extinction*. We have that

$$1 - p_0(t) = 1 - \mu\beta(t) = \frac{(\mu - \lambda)e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}. \quad (5.40)$$

¹⁰Conny Palm (1907-1951), mathematician and telecommunications engineer in Stockholm.

Therefore, if $\mu \neq \lambda$,

$$P(0, t) := \frac{(\mu - \lambda)e^{(\lambda - \mu)t}}{\mu - \lambda e^{(\lambda - \mu)t}} \quad (5.41)$$

is the probability that a Markov branching process, which starts with one lineage has some descendant at time t . Note that $P(0, 0) = 1$. We can also write

$$p_1(t) = P(0, t) \cdot (1 - \lambda\beta(t)), \quad t > 0. \quad (5.42)$$

Let us next assume

$$\frac{\lambda}{\mu} < 1. \quad (5.43)$$

Under the assumption (5.43) it holds obviously that

$$P(0, t) \rightarrow 0, \quad \text{as } t \rightarrow \infty,$$

and thus it holds for the probability of extinction that

$$p_0(t) \rightarrow 1, \quad \text{as } t \rightarrow \infty. \quad (5.44)$$

This is seen as almost certain extinction (in finite time). As an identity it holds also that

$$\lambda\beta(t) = 1 - \frac{(\mu - \lambda)}{\mu - \lambda e^{(\lambda - \mu)t}}. \quad (5.45)$$

We see now under (5.43) that

$$\lambda\beta(t) \rightarrow \frac{\lambda}{\mu}, \quad \text{as } t \rightarrow \infty. \quad (5.46)$$

Next we observe the presence of a standard geometric distribution in this model. This follows directly from (5.36), (5.38), and by the definition of a conditional probability.

Lemma 5.1 *The probability of n lineages conditioned on the event that the birth-death process is not extinct at time t is*

$$p_n^{(*)}(t) = (1 - \lambda\beta(t)) (\lambda\beta(t))^{n-1}, \quad t > 0, \quad n \geq 1. \quad (5.47)$$

■

We get also from (5.46) and (5.32) that

$$p_n^{(*)}(t) \rightarrow \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{n-1} \quad n \geq 1. \quad (5.48)$$

This is in other words the (quasi)stationary probability distribution for the number of lineages under the condition that the process has not gone extinct.

Example 5.2 (A Pure Birth Process) The pure birth process, also known as a *Yule process*, is obtained from the preceding by setting $\mu = 0$ in (5.30) and this gives

$$p_1'(t) = -\lambda p_1(t) \tag{5.49}$$

$$p_j'(t) = (j-1)\lambda p_{j-1}(t) - j\lambda p_j(t), \quad j > 1$$

with

$$p_1(0) = 1, \quad p_j(0) = 0, \quad \text{for } j \neq 1.$$

Then we get with (5.33)

$$\lambda\beta(t) = 1 - e^{-\lambda t}, \tag{5.50}$$

and

$$p_n(t) = e^{-\lambda t} \cdot (1 - e^{-\lambda t})^{n-1}, \quad t > 0, \quad n \geq 1. \tag{5.51}$$

■

6 Power Laws by Yule, and Reed & Hughes

6.1 Birth and Death Processes

Next we show how power laws are derived from the expressions for $p_n(t)$ for birth-and-death processes in the preceding section. The analysis is taken from [22, 23]. The case considered by Yule in [31] is given a special treatment in the next subsection.

We assume that $T \in \text{Exp}(\rho)$ independent of the birth-and-death process $X(t)$ and consider

$$q_{n+1} = P(X(T) = n + 1), \quad n \geq 1,$$

which is the birth-and-death process killed or observed at T . The notion behind T is that at the time we observe the various domains, these have lived different times, which we assume to be exponentially distributed.

A standard rule of probability calculus gives that

$$q_{n+1} = P(X(T) = n + 1) = \int_0^\infty P(X(t) = n + 1 \mid T = t) f_T(t) dt$$

$$= \int_0^\infty P(X(t) = n+1) f_T(t) dt,$$

since $X(t)$ is independent of T . By assumption $f_T(t)$ is the density of $\text{Exp}(\rho)$, and thus we have

$$= \int_0^\infty p_{n+1}(t) \rho e^{-\rho t} dt.$$

We insert from (5.39) and get

$$q_{n+1} = \int_0^\infty \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)t}}{(\lambda - \mu e^{-(\lambda - \mu)t})^2} \left[\frac{\lambda e^{-(\lambda - \mu)t} - \lambda}{\mu e^{-(\lambda - \mu)t} - \lambda} \right]^n \rho e^{-\rho t} dt \quad (6.52)$$

$$= \int_0^\infty \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)t}}{(\lambda - \mu e^{-(\lambda - \mu)t})^2} \left[\frac{\lambda - \lambda e^{-(\lambda - \mu)t}}{\lambda - \mu e^{-(\lambda - \mu)t}} \right]^n \rho e^{-\rho t} dt. \quad (6.53)$$

We make the change of variable of integration from t to τ by

$$t = (\lambda - \mu)^{-1} \log [n(1 - \mu/\lambda) / \tau].$$

Thus $t = \infty \leftrightarrow \tau = 0$, $t = 0, \leftrightarrow \tau = n(1 - \mu/\lambda)$, and

$$dt = \frac{-1}{(\lambda - \mu)\tau} d\tau.$$

In addition

$$e^{-(\lambda - \mu)t} = \frac{\lambda}{\lambda - \mu} \frac{\tau}{n},$$

$$\lambda - \mu e^{-(\lambda - \mu)t} = \lambda - \mu \frac{\lambda}{\lambda - \mu} \frac{\tau}{n} \approx \lambda,$$

for large n ,

$$e^{-\rho t} = (1 - 1 - \mu/\lambda)^{-\rho/(\lambda - \mu)} \cdot \tau^{-\rho/(\lambda - \mu)} \cdot n^{-\rho/(\lambda - \mu)},$$

and

$$\begin{aligned} \left[\frac{\lambda - \lambda e^{-(\lambda - \mu)t}}{\lambda - \mu e^{-(\lambda - \mu)t}} \right]^n &= \left[\frac{1 - \frac{\lambda}{\lambda - \mu} \frac{\tau}{n}}{1 - \frac{\mu}{\lambda - \mu} \frac{\tau}{n}} \right]^n \\ &\sim \frac{e^{-\frac{\lambda}{\lambda - \mu} \tau}}{e^{-\frac{\mu}{\lambda - \mu} \tau}} = e^{-\tau} \end{aligned}$$

Collecting from the above we get after some algebra that

$$q_{n+1} \sim \frac{\rho}{\lambda} (1 - \mu/\lambda)^{-\rho/(\lambda - \mu)} \cdot n^{-1 - \rho/(\lambda - \mu)} \int_0^\infty e^{-\tau} \tau^{-\rho/(\lambda - \mu)} d\tau.$$

We note that

$$\Gamma(1 + \rho/(\lambda - \mu)) = \int_0^\infty e^{-\tau} \tau^{-\rho/(\lambda - \mu)} d\tau.$$

Hence we have for large n the power law with $\gamma = 1 + \rho/(\lambda - \mu)$,

$$q_n \sim C \cdot n^{-1 - \rho/(\lambda - \mu)}, \quad (6.54)$$

where

$$C = \frac{\rho}{\lambda} \cdot (1 - \mu/\lambda)^{-\rho/(\lambda - \mu)} \cdot \Gamma(1 + \rho/(\lambda - \mu)).$$

6.2 Pure Birth Processes and the Yule Distribution

We consider example 5.2 and the result in (5.51). We have

$$p_n(t) = \text{the relative frequency of a domain family of size } n = P(X(t) = n),$$

Then, we have obtained in (5.51) that

$$p_n(t) = e^{-\lambda t} \cdot (1 - e^{-\lambda t})^{n-1}, \quad t > 0, \quad n \geq 1. \quad (6.55)$$

Let now T be a random variable with exponential distribution $\text{Exp}(\rho)$ and independent of $X(t)$. We want to find

$$q_n = P(X(T) = n), \quad n \geq 1,$$

which is the pure birth process killed or observed at T . It holds by a standard rule of probability calculus that

$$\begin{aligned} q_n &= P(X(T) = n) = \int_0^\infty P(X(t) = n \mid T = t) f_T(t) dt \\ &= \int_0^\infty P(X(t) = n) f_T(t) dt, \end{aligned}$$

since $X(t)$ is independent of T . By assumption $f_T(t)$ is the density of $\text{Exp}(\rho)$, and thus we have

$$= \int_0^\infty p_n(t) \rho e^{-\rho t} dt.$$

Then we substitute from (6.55), and find

$$q_n = \rho \int_0^\infty e^{-\rho t} e^{-\lambda t} \cdot (1 - e^{-\lambda t})^{n-1} dt$$

$$= \rho \int_0^\infty (e^{-\lambda t})^{\frac{\rho}{\lambda}} e^{-\lambda t} \cdot (1 - e^{-\lambda t})^{n-1} dt.$$

The purpose of the strange-looking way of rewriting $e^{-\rho t}$ will become clear, when we make the change of variable

$$u = e^{-\lambda t}, \quad dt = \frac{-1}{\lambda \cdot u} du, t = 0 \leftrightarrow u = 1, t = \infty \leftrightarrow u = 0,$$

since this gives us

$$\begin{aligned} &= -\rho \int_1^0 (u)^{\frac{\rho}{\lambda}} u \cdot (1 - u)^{n-1} \frac{1}{\lambda \cdot u} du \\ &= \frac{\rho}{\lambda} \int_0^1 (u)^{\frac{\rho}{\lambda}} \cdot (1 - u)^{n-1} du. \end{aligned}$$

Here we recognize the Beta integral of (1.7), and get

$$q_n = \frac{\rho}{\lambda} B\left(\frac{\rho}{\lambda} + 1, n\right), \quad n \geq 1. \quad (6.56)$$

This is the distribution in (1.6) with $\delta = \frac{\rho}{\lambda}$ and $k = n$. The power-law tail

$$q_n \sim c \cdot n^{-1-\rho/\lambda} \quad (6.57)$$

is established in (B.5) in the appendix B. This is also obtained, as it should, by setting $\mu = 0$ in (6.54).

The constant $\delta = \frac{\rho}{\lambda}$ is seen to have a biological interpretation. We can take ρ as the rate of mutation at generic level, and λ as the rate of mutation at species level.

A Notations

The notation $f(x) \sim g(x)$ (at $x = a$) means [5, p. 432] the following

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1. \quad (A.1)$$

This means that the functions grow at the same rate at a . For example, if

$$f(x) = x^2, g(x) = x^2 + x,$$

then

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{1}{1 + \frac{1}{x}} = 1,$$

but at the same time $f(x) - g(x) = x$.

B The Yule-Simon Distribution has a Power-Law Tail

We want to establish that

$$P(X = k) = \rho B(k, \rho + 1), k = 1, 2, \dots, \quad (\text{B.1})$$

is a power law. First we write $B(x, y)$ using the beta integral, or in other words

$$B(x, y) = \int_0^1 u^{x-1} \cdot (1-u)^{y-1} du.$$

We assume $y > 1$. The argument below follows in the main the treatment in [29, p. 58].

Let us first substitute $u = e^{-t}$, and get

$$\begin{aligned} \int_0^1 u^{x-1} \cdot (1-u)^{y-1} du &= - \int_\infty^0 e^{-t(x-1)} \cdot (1-e^{-t})^{y-1} e^{-t} dt \\ &= \int_0^\infty e^{-tx} \cdot (1-e^{-t})^{y-1} dt = I_1 - I_2, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int_0^\infty e^{-tx} \cdot t^{y-1} dt, \\ I_2 &= \int_0^\infty \left(t^{y-1} - (1-e^{-t})^{y-1} \right) \cdot e^{-tx} dt. \end{aligned}$$

In the integral I_1 we make the change of variable $u = xt$, $dt = \frac{1}{x} du$, and get

$$\begin{aligned} \int_0^\infty e^{-tx} \cdot t^{y-1} dt &= \int_0^\infty e^{-u} \cdot \frac{u^{y-1}}{x^{y-1}} \frac{1}{x} du \\ &= \frac{1}{x^y} \int_0^\infty e^{-u} \cdot u^{y-1} du = \frac{\Gamma(y)}{x^y}, \end{aligned} \quad (\text{B.2})$$

where we applied (1.8). Thus we have

$$B(x, y) = \frac{\Gamma(y)}{x^y} - \int_0^\infty \left(t^{y-1} - (1-e^{-t})^{y-1} \right) \cdot e^{-tx} dt. \quad (\text{B.3})$$

We find an upper bound for the integral I_2 in the right hand side of (B.3). We note that $I_2 > 0$, because $1 - e^{-t} < t$, and

$$I_2 = \int_0^\infty \left(t^{y-1} - (1-e^{-t})^{y-1} \right) \cdot e^{-tx} dt$$

$$\leq \int_0^1 \left(t^{y-1} - (1 - e^{-t})^{y-1} \right) \cdot e^{-tx} dt + \int_1^\infty t^{y-1} \cdot e^{-tx} dt.$$

For $0 \leq t \leq 1$, it holds that

$$1 - e^{-t} \geq t - \frac{1}{2}t^2,$$

and therefore

$$\int_0^1 \left(t^{y-1} - (1 - e^{-t})^{y-1} \right) \cdot e^{-tx} dt \leq \int_0^1 \left(t^{y-1} - \left(t - \frac{1}{2}t^2 \right)^{y-1} \right) \cdot e^{-tx} dt$$

and

$$\begin{aligned} & \int_0^1 \left(t^{y-1} - \left(t - \frac{1}{2}t^2 \right)^{y-1} \right) \cdot e^{-tx} dt = \\ & = \int_0^1 \left(1 - \left(1 - \frac{1}{2}t \right)^{y-1} \right) \cdot t^{y-1} e^{-tx} dt \leq K(y) \int_0^1 t^{y-1} e^{-tx} dt. \end{aligned}$$

where $K(y) = (1 - 2^{-(y-1)})$. Thus we have obtained the bound

$$\begin{aligned} I_2 &= \int_0^\infty \left(t^{y-1} - (1 - e^{-t})^{y-1} \right) \cdot e^{-tx} dt \\ &\leq K(y) \int_0^1 t^{y-1} e^{-tx} dt + \int_1^\infty t^{y-1} \cdot e^{-tx} dt. \end{aligned}$$

For $t \geq 1$, and $y > 1$, we have $t^{y-1} \leq t^y$, which gives

$$\begin{aligned} I_2 &\leq K(y) \int_0^1 t^{y-1} e^{-tx} dt + \int_1^\infty t^y \cdot e^{-tx} dt \\ &\leq K(y) \int_0^1 t^{y-1} e^{-tx} dt + \int_0^\infty t^y \cdot e^{-tx} dt = \\ &= K(y) \int_0^1 t^{y-1} e^{-tx} dt + \frac{\Gamma(y+1)}{x^{y+1}}, \end{aligned}$$

where we reinvoked the change of variable ($u = xt$) from the above. By the same change of variable

$$\int_0^1 t^{y-1} e^{-tx} dt = \frac{1}{x^y} \int_0^x u^{y-1} e^{-u} du = H(x).$$

Thus we have the bound

$$I_2 \leq \frac{K(y)}{x^y} H(x) + \frac{\Gamma(y+1)}{x^{y+1}}. \quad (\text{B.4})$$

In other words, we have by (B.3) and (B.4) obtained

$$B(x, y) = I_1 - I_2 \geq \frac{\Gamma(y)}{x^y} - \frac{K(y)}{x^y} H(x) - \frac{\Gamma(y+1)}{x^{y+1}}$$

Also, as $I_2 > 0$,

$$B(x, y) = I_1 - I_2 < I_1 = \frac{\Gamma(y)}{x^y}.$$

Hence

$$\frac{\Gamma(y)}{x^y} - \frac{K(y)}{x^y} H(x) - \frac{\Gamma(y+1)}{x^{y+1}} \leq B(x, y) \leq \frac{\Gamma(y)}{x^y}.$$

By definition of the beta function $B(x, y)$ we have

$$\frac{\Gamma(y)}{x^y} - \frac{K(y)}{x^y} H(x) - \frac{\Gamma(y+1)}{x^{y+1}} \leq \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \leq \frac{\Gamma(y)}{x^y}$$

Thus

$$1 - \frac{K(y)}{\Gamma(y)} H(x) - \frac{\Gamma(y+1)/\Gamma(y)}{x} \leq \frac{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}}{\frac{\Gamma(y)}{x^y}} \leq 1$$

As $x \rightarrow \infty$, $\int_0^x u^{y-1} e^{-u} du \rightarrow \int_0^\infty u^{y-1} e^{-u} du = \Gamma(y)$, and thus

$$\frac{K(y)}{\Gamma(y)} H(x) \rightarrow \frac{K(y)}{\Gamma(y)} \Gamma(y) = K(y).$$

Hence we have obtained for $y > 1$ that

$$1 - K(y) \leq \lim_{x \rightarrow \infty} \left[\frac{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}}{\frac{\Gamma(y)}{x^y}} \right] \leq 1$$

Since $K(y) = \left(1 - 2^{-(y-1)}\right)$ can be arbitrarily close to zero independently of x , we have for all practical purposes established that

$$\frac{\Gamma(x)}{\Gamma(x+y)} \sim \frac{1}{x^y}, \quad \text{as } x \rightarrow \infty. \quad (\text{B.5})$$

Let us note that this constitutes a liberal use of \sim as defined in (A.1). The case $y \leq 1$ can be treated using the argument above and the properties of the Euler function. We have thus shown that the Yule-Simon Distribution is a power law, as claimed in (1.9).

C Probability Generating Functions

The method of solution (which, as we have claimed, is due to Conny Palm) of the equations (5.30), or of

$$p_0'(t) = \mu p_1(t) \tag{C.1}$$

$$p_j'(t) = (j-1)\lambda p_{j-1}(t) - j(\lambda + \mu)p_j(t) + (j+1)\mu p_{j+1}(t), \quad j \geq 1.$$

is based on a probability generating function. By definition (see [5, p.19]) the probability generating function is

$$G(t, s) = \sum_{j=0}^{\infty} s^j p_j(t).$$

From the initial condition in (5.31) we get obviously

$$G(0, s) = s.$$

On the other hand, we can multiply in (C.1) by s^j and sum over j to get

$$\sum_{j=0}^{\infty} s^j p_j'(t) = \lambda \sum_{j=1}^{\infty} (j-1) s^j p_{j-1}(t) - (\lambda + \mu) \sum_{j=0}^{\infty} j s^j p_j(t) + \mu \sum_{j=0}^{\infty} (j+1) s^j p_{j+1}(t). \tag{C.2}$$

We observe that

$$\frac{\partial}{\partial t} G(t, s) = \sum_{j=0}^{\infty} s^j p_j'(t).$$

Then we get in (C.2) that

$$\frac{\partial}{\partial t} G(t, s) = \lambda s^2 \sum_{j=1}^{\infty} (j-1) s^{j-2} p_{j-1}(t) - (\lambda + \mu) s \sum_{j=0}^{\infty} j s^{j-1} p_j(t) + \mu \sum_{j=0}^{\infty} (j+1) s^j p_{j+1}(t).$$

We observe

$$\frac{\partial}{\partial s} G(t, s) = \sum_{j=1}^{\infty} j s^{j-1} p_j(t),$$

and by a change of variable of summation $j-1 = h$ and then setting again $j = h$, that

$$\sum_{j=1}^{\infty} (j-1) s^{j-2} p_{j-1}(t) = \sum_{h=0}^{\infty} h s^{h-1} p_h(t) = \sum_{j=0}^{\infty} j s^{j-1} p_j(t) = \frac{\partial}{\partial s} G(t, s),$$

as well as (change of variable of summation $j + 1 = h$)

$$\sum_{j=0}^{\infty} (j+1)s^j p_{j+1}(t) = \frac{\partial}{\partial s} G(t, s).$$

Thus we have

$$\frac{\partial}{\partial t} G(t, s) = \lambda s^2 \frac{\partial}{\partial s} G(t, s) - (\lambda + \mu)s \frac{\partial}{\partial s} G(t, s) + \mu \frac{\partial}{\partial s} G(t, s), \quad (\text{C.3})$$

which gives

$$\frac{\partial}{\partial t} G(t, s) = (\lambda s - \mu)(s - 1) \frac{\partial}{\partial s} G(t, s) \quad (\text{C.4})$$

with the boundary condition

$$G(0, s) = s. \quad (\text{C.5})$$

This is a linear first order partial differential equation that can be solved by standard methods (of characteristic curves)¹¹. We have for $\mu \neq \lambda$

$$G(t, s) = \left(\frac{\mu \cdot (1 - s) - (\mu - \lambda)s e^{-t(\lambda - \mu)}}{\lambda \cdot (1 - s) - (\mu - \lambda)s e^{-t(\lambda - \mu)}} \right). \quad (\text{C.6})$$

Then, as is known [5, p.20], we can obtain $p_n(t)$ in (5.39) by differentiation

$$\frac{\partial^n}{\partial s^n} G(t, s) \Big|_{s=0} = p_n(t),$$

as may be verified.

Since expectation and variance are also derivable from a probability generating function, we get

$$\frac{\partial}{\partial s} G(t, s) \Big|_{s=1} = E[X(t)] = e^{(\mu - \lambda)t}, \quad (\text{C.7})$$

and, for $\lambda \neq \mu$, by [5, p. 3]

$$\begin{aligned} \frac{\partial^2}{\partial s^2} G(t, s) \Big|_{s=1} + E[X(t)] - [E[X(t)]]^2 &= \text{Var}[X(t)] \\ &= \frac{\lambda + \mu}{\lambda - \mu} e^{(\mu - \lambda)t} (e^{(\mu - \lambda)t} - 1). \end{aligned} \quad (\text{C.8})$$

¹¹Metoden med karakteristikerna framställs t.ex. i kapitlen 1–2 av J. Malmquist, V. Stenström, S. Danielson: *Matematisk analys. Del III*, Natur och Kultur, Uppsala 1953.

D Weak Versions of Zipf's Law

Here we recapitulate without proofs the results due to Hill [7]. These results have been refined and extended in [8, 9]. We suppose that a population consists of N species and M non-empty genera. We let L_i the number of species belonging to the i th genus, $i = 1, 2, \dots, M$, and then

$$\sum_{i=1}^M L_i = N. \quad (\text{D.1})$$

We set

$$\mathbf{L} = (L_1, \dots, L_M).$$

We assume that

$$P(\mathbf{L} \mid N, M) = \frac{1}{\binom{N-1}{M-1}}. \quad (\text{D.2})$$

This can be viewed as a discrete analogue of the M -dimensional, degenerate, uniform distribution for the proportions $N^{-1}L_i$ on the M -dimensional simplex. The random allocation procedure (the *Bose-Einstein procedure*) yielding (D.2) is that of placing N undistinguishable balls in M urns. In order to secure that all M urns are non-empty, we first put one ball in each of the urns, and then distribute the remaining $N - M$ balls in the M urns at random. The number of ways of doing this is

$$\binom{M-1+(N-M)}{N-M} = \binom{N-1}{M-1}.$$

Let next a_s be the (random) number of genera with exactly s species. Hence, if

$$\begin{aligned} S &= N - M + 1, \\ \sum_{s=1}^S sa_s &= M. \end{aligned}$$

Then, the size distribution is

$$\mathbf{a} = (a_1, \dots, a_N),$$

we have

$$P(\mathbf{a} \mid N, M) = \frac{M!}{\prod_{s=1}^S a_s!} \frac{1}{\binom{N-1}{M-1}}.$$

For this we note that the same \mathbf{a} will arise from any permutation of the components of the vector \mathbf{L} , which has exactly a_s components of with the value s , and that there are $\frac{M!}{\prod_{s=1}^s a_s!}$ combinations of giving rise to distinct vectors \mathbf{L} with such components.

Then

$$\frac{1}{M}a_s$$

is the observed relative frequency of the number of genera with s species.

Hill shows that, when $M, N \rightarrow \infty$, so that $\frac{M}{N} \rightarrow \theta$, there is convergence in probability

$$\frac{1}{M}a_s \xrightarrow{p} h_s(\theta),$$

where we have introduced

$$h_s(\theta) = \theta \cdot (1 - \theta)^{s-1}, \quad (\text{D.3})$$

where θ is a random variable with the distribution $F(\cdot)$, obtained as

$$P\left(\frac{M}{N} \leq x \mid N\right) \rightarrow F(x),$$

In the limit we have also

$$E\left[\frac{1}{M}a_s \mid N\right] \rightarrow E[h_s(\theta)] = \int_0^1 \theta \cdot (1 - \theta)^{s-1} dF(\theta).$$

Hence, if the distribution of θ is the Beta distribution with parameters α and β , we use (1.7) to get

$$\begin{aligned} E[h_s(\theta)] &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta \cdot (1 - \theta)^{s-1} \theta^{\alpha-1} \cdot (1 - \theta)^{\beta-1} d\theta \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 1)\Gamma(\beta + s - 1)}{\Gamma(s + \alpha + \beta)}. \end{aligned} \quad (\text{D.4})$$

With $\alpha = 1$, $\beta = \delta$ this coincides with the Yule-Simon law in (1.6).

If θ has the uniform distribution (=the Beta distribution with parameters $\alpha = \beta = 1$), the computation above gives

$$E[h_s(\theta)] = \frac{1}{s(s-1)}, \quad (\text{D.5})$$

which is a particularly simple form of Zipf's law.

The results in (D.4) and (D.4) hold for expected size distributions, and are thus **weak versions of Zipf's law**.

Remark D.1 The function in (D.3) appears in the **Griffiths-Engen-McCloskey** (GEM) distribution, see e.g., [4, 10]. Let $\{P_n\}_{n \geq 1}$ be a sequence of independent, identically Beta distributed, with parameters $\alpha = 1$ and β , random variables. We consider the sequence

$$Q_1 = P_1, Q_i = P_i \cdot \prod_{j=1}^{i-1} (1 - P_j), \quad i = 2, 3, \dots$$

The distribution of

$$Q^* = \{Q_i\}_{i \geq 1}$$

is the GEM distribution¹².

By independence we get

$$E[Q_i] = E[P_i] \cdot \prod_{j=1}^{i-1} E[(1 - P_j)],$$

where (1.7) gives

$$\begin{aligned} E[P_i] &= \int_0^1 x \cdot \beta(1-x)^{\beta-1} dx = \frac{\beta\Gamma(2)\Gamma(\beta)}{\Gamma(\beta+2)} \\ &= \frac{1}{(\beta+1)}, \end{aligned}$$

where we used $\Gamma(z+1) = z\Gamma(z)$. In the same way we get

$$E[(1 - P_j)] = \int_0^1 \beta(1-x)^\beta dx = \frac{\beta}{(\beta+1)}.$$

If we set

$$\theta = \frac{1}{(\beta+1)},$$

then we have in view of the notation in (D.3) obtained

$$E[Q_i] = \frac{1}{(\beta+1)} \left(\frac{\beta}{(\beta+1)} \right)^{i-1} = h_i(\theta).$$

¹² Q^* is a size-biased permutation of a random probability vector, which is Poisson-Dirchlet distributed, [10].

References

- [1] S.P. Buldyrev, A.L. Goldberger, S. Havlin, C-K. Peng, M. Simons, and H.E. Stanley (1993): Generalized Lévy Random walk model for DNA nucleotide sequences. *Physical Review E*, 47, 4513–4524.
- [2] D.Y.C. Chan, B.D. Hughes, and A.S. Leong (2003): Stochastically evolving networks. *Physical Review E*, 68, 066124.
- [3] S.N. Dorogovtsev, and J.F.F. Mendes (2002): Evolution of Networks. *Advances on Physics*, 51, pp. 1079–1187.
- [4] S. Engen (1975): A Note on the Geometric Series as a Species Frequency Model. *Biometrika*, 62, pp. 697–699.
- [5] W.J. Ewens and G.R. Grant (2001): *Statistical Methods in Bioinformatics. An Introduction*. Springer, New York. Berlin, Heidelberg, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokyo.
- [6] G.R. Grimmett and D.R. Stirzaker (1994): *Probability and Random Processes. A Second Edition*. Oxford Science Publications, Oxford.
- [7] B.M. Hill (1970): Zipf’s Law and Prior Distribution for the Composition of a Population. *Journal of the American Statistical Association*, 65, pp. 1220–1232.
- [8] B.M. Hill (1974): The Rank-frequency Form of Zipf’s Law. *Journal of the American Statistical Association*, 69, pp. 1017–1026.
- [9] B.M. Hill, and M. Woodroffe (1975): Stronger Forms of Zipf’s Law. *Journal of the American Statistical Association*, 70, pp. 212–219.
- [10] L. Holst (2001): The Poisson-Dirchlet Distribution and its Relatives Revisited. *preprint*
http://www.math.kth.se/gunnare/lasse_s.htm
- [11] M.A. Huynen, and E. van Nimwegen (1998): The Frequency Distribution of Gene Family Sizes in Complete Genomes. *Genome Biology*
<http://www.genomebiology.com/2002/3/8/research/0040.1>

- [12] G.P. Karev, Y.I. Wolf, A.Y. Rzhetsky, F.S. Berezovskaya, and E.P. Koonin (2002): A Simple Model of Evolution Explains Power Law Behaviour. *BMC Evolutionary Biology*, 2:18
<http://www.biomedcentral.com/1471-2148/2/18>
- [13] G.P. Karev, Y.I. Wolf, F.S. Berezovskaya, and E.P. Koonin (2003): Simple stochastic birth and death models of genome evolution: was there enough time for us to evolve ? *Bioinformatics*, 19, pp. 1898–1900.
- [14] G.P. Karev, Y.I. Wolf, F.S. Berezovskaya, and E.P. Koonin (2005): Gene family evolution: an in-depth theoretical and simulation analysis of non-linear birth-death-innovation models. *BMC Evolutionary Biology*, 4:42
<http://www.biomedcentral.com/1471-2148/4/32>
- [15] H. Kesten (1973): Random Difference Equations and Renewal Theory for Products of Random Matrices. *Acta Mathematica*, 1973, pp. 207–248.
- [16] T. Koski (2004): Lectures at RNI on Probabilistic Models and Inference for Phylogenetics. *Linköping Studies in mathematics* 2004:001, Linköping University Electronic Press,
<http://www.ep.liu.se/ea/lsm/2004/001/>.
- [17] V.A. Kuznetsov (2001): Distribution Associated with Stochastic Processes of Gene Expression in a Single Eukaryotic Cell. *EURASIP Journal of Applied Signal Processing*, 4, pp. 285–296.
- [18] F. Liljeroos, C.F.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Åberg (2001): The web of human sexual contacts. *Nature*, 411, pp. 907–908.
- [19] <http://www.nist.gov/dads/HTML/lotkaslaw.html>
- [20] N.M. Luscombe, J. Qian, Z. Zhang, T. Johnson and M. Gerstein (2002): The dominance of the population by the few: power-law behaviour applies to a wide variety of genomic properties. *Molecular Biology and Evolution*, 15, pp. 583–589.
- [21] J.J. Ramsden, and J. Vohradsky (1998): Zipf-like behavior in prokaryotic protein expression. *Physical Review E*, 58, pp. 7777–7780.

- [22] W.J. Reed, and B.D. Hughes (2002): On the Size Distribution of Live Genera. *Journal of Theoretical Biology*, 217, pp. 125–135.
- [23] W.J. Reed, and B.D. Hughes (2002): From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E*, 66, pp. 067103.
- [24] W.J. Reed, and B.D. Hughes (2004): A model explaining the size distribution of gene and protein families. *Mathematical Biosciences*, 19, pp. 97–102.
- [25] A. Rouault (1978): Lois de Zipf et sources markoviennes. *Annales de l'institut Henri Poincaré-Section B*, XIV, pp. 1694–188.
- [26] A. Rzhetsky, and S.M Gomez (2001): Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome. *Bioinformatics*, 17, pp. 98–996.
- [27] H.A. Simon (1955): On a Class of Skew Distribution Functions. *Biometrika*, 42, 3/4, pp. 425–440.
- [28] D. Sornette, and Th.M. Niewenhuisen (1997): Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws. *Journal de Physique I*, 7, 3, pp. 431–444.
- [29] E.C. Titchmarsh (1939): *The Theory of Functions. Second Edition*. Oxford University Press, London.
- [30] G Troll, and P. beim Graben (1998): Zipf's law is not a consequence of the central limit theorem. *Physical Review E*, 57, pp. 1347–1355.
- [31] G.U. Yule (1925): A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.. *Philosophical Transactions Royal Society B*, 213, pp. 431–444.